



Platform Linked
Data Nederland



Inloop introductie

LINKED DATA VOOR DE LIEFHEBBER

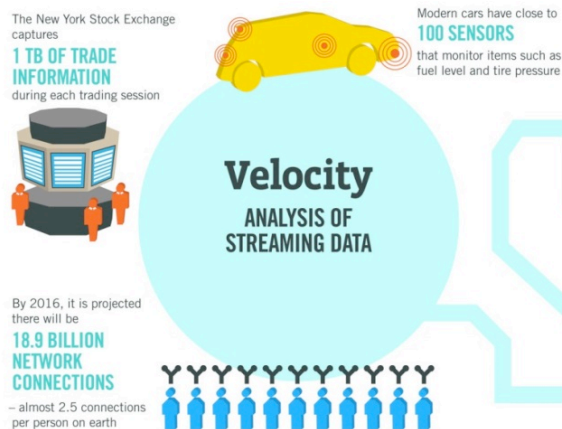
water**net**







Platform Linked
Data Nederland



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
(161 BILLION GIGABYTES)

Variety DIFFERENT FORMS OF DATA

**30 BILLION
PIECES OF CONTENT**
are shared on Facebook every month

**4 BILLION+
HOURS OF VIDEO**
are watched on YouTube each month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

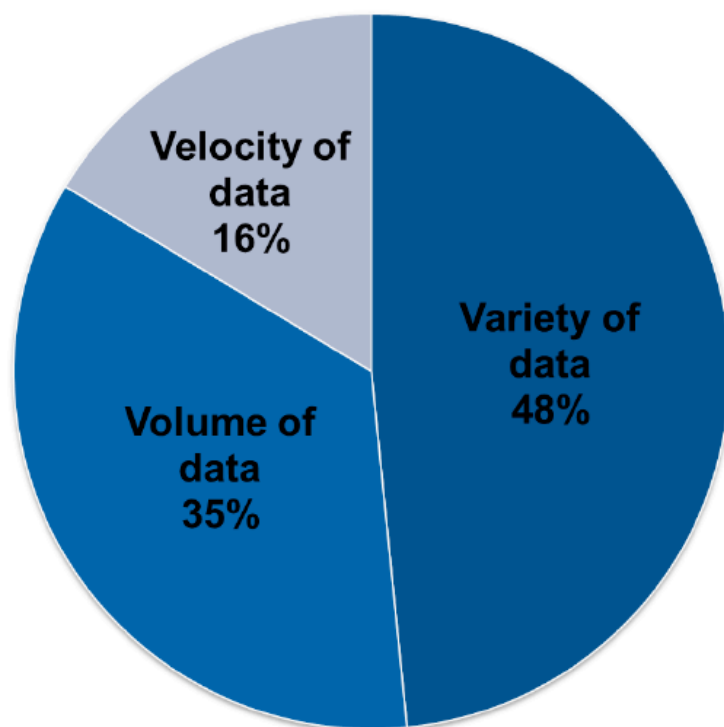
Veracity UNCERTAINTY OF DATA

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS

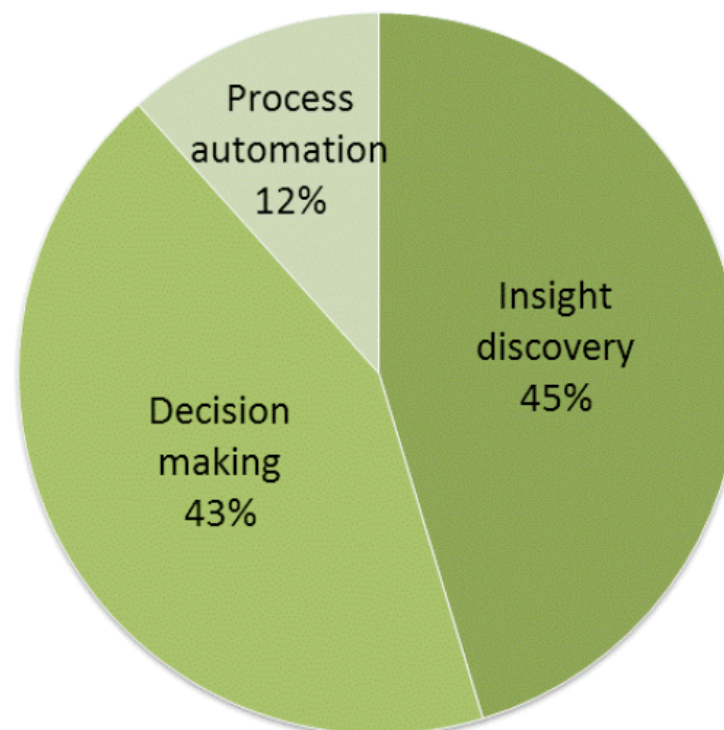
IBM

Acknowledge big data initiatives are unique

Biggest Big Data Challenge



Biggest Big Data Opportunity

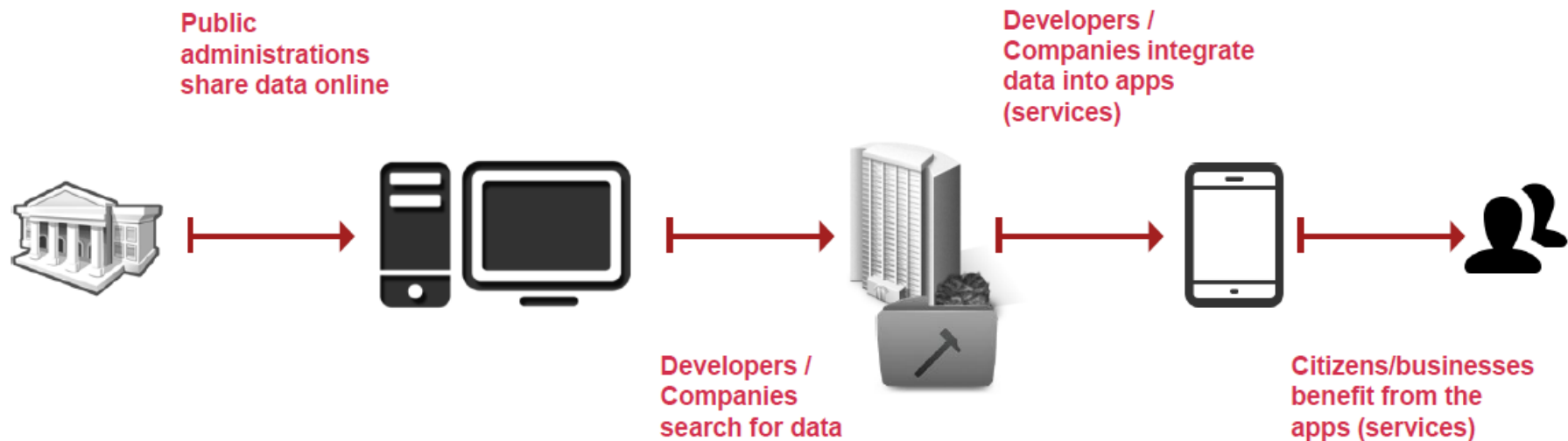


Gartner: Big Data Makes Firms Smarter; Open Data Makes Them Richer



Open Government Data has a great potential to create social and economic value

Publishing data

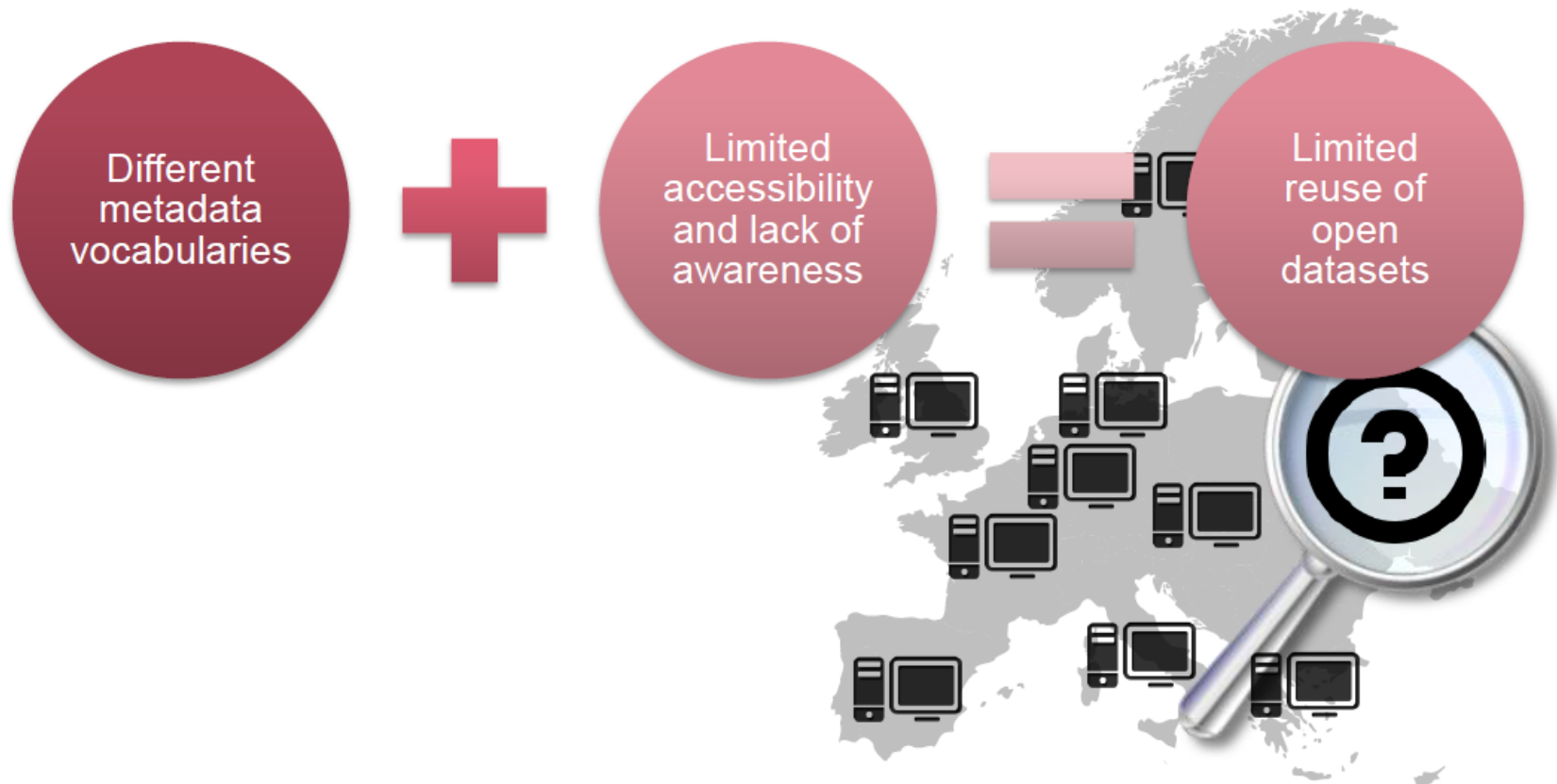


Reusing data

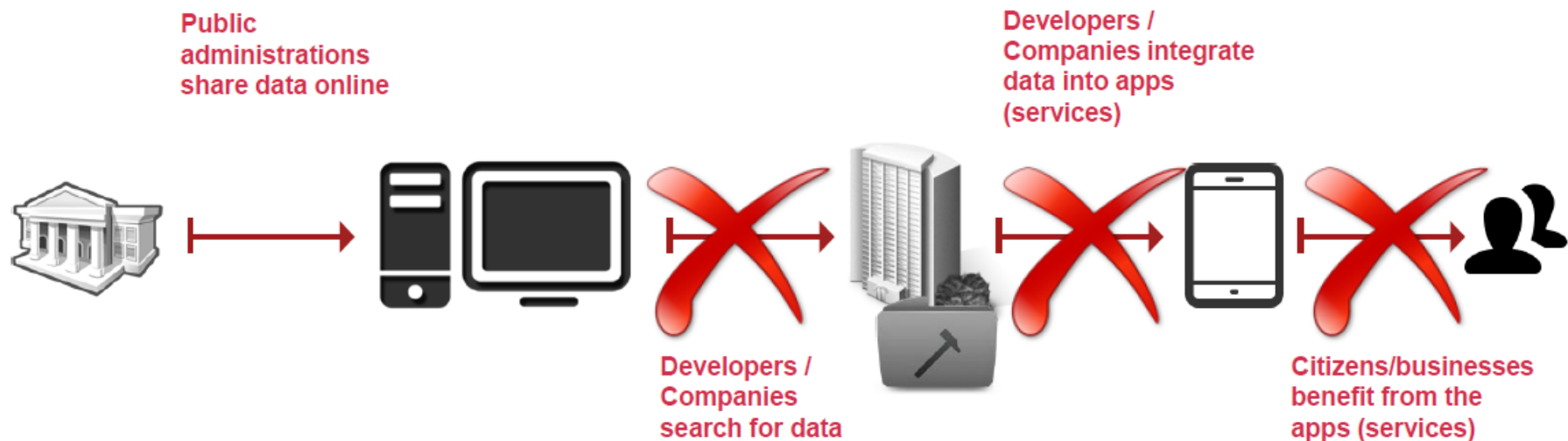
Barriers to Open Data publishing and reuse

		Data publishers	Data reusers		
		No view on which data is more likely to be reused / has a higher ROI potential.	Lack of overview of existing/available datasets.		
		Unclear business model for publishing Open Data.	Unclear business model for reusing Open Data.		
		Limited tool support.	Data is often of low quality, outdated, unstructured and/or not machine-readable.		
		Competing licences for datasets.	Lack of licensing information or incompatible licences.		
Metadata	{	Competing vocabularies for describing datasets.	Different vocabularies when searching for datasets.	{	Metadata
		Domain-specific metadata needs.	Lack of (good quality) metadata.		
		Effort required for keeping the metadata up-to-date.	Lack of provenance information.		

Limited accessibility and lack of (cross-border/sector) awareness of open datasets



No reuse = No social and economic value





Platform Linked
Data Nederland

Kan open, kan big, kan smart.....maar belangrijkste is dat data herbruikbaar is!

WE WILLEN BRILJANTE DINGEN MET DATA KUNNEN DOEN!

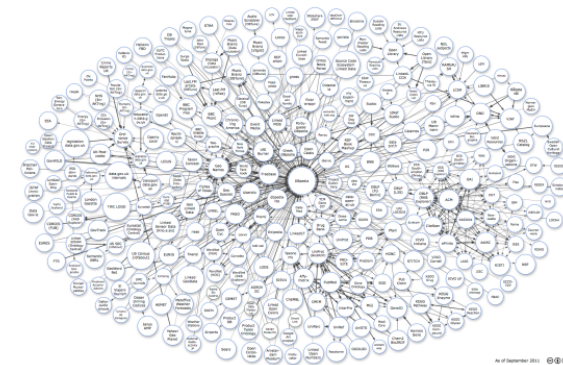
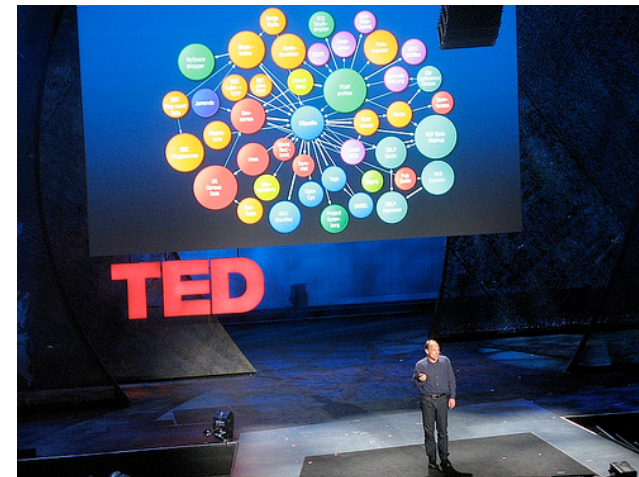






Platform Linked
Data Nederland

Waar komt het vandaan?





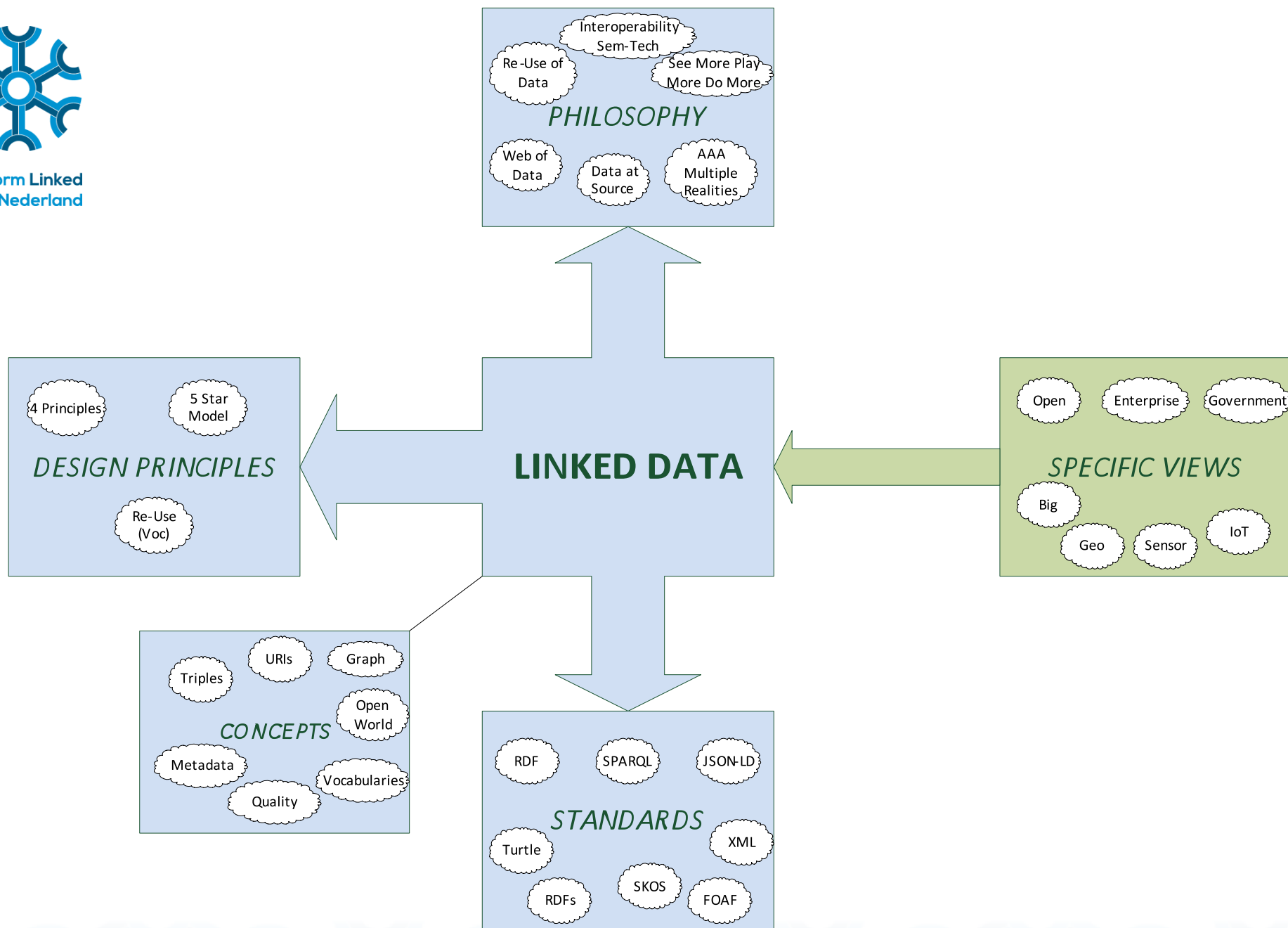
Platform Linked
Data Nederland

Maar wat is het?





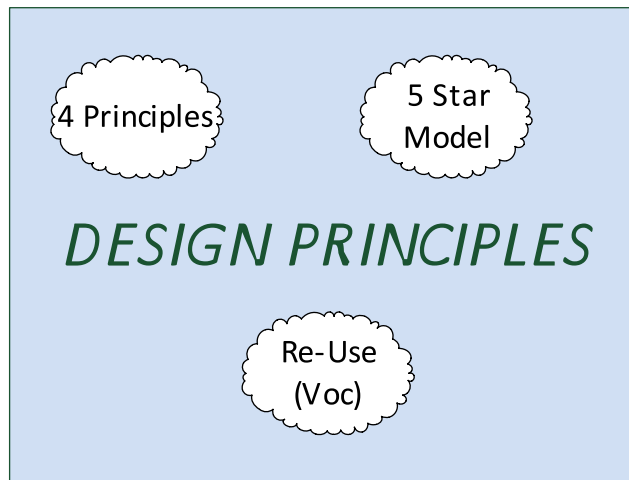
Platform Linked
Data Nederland



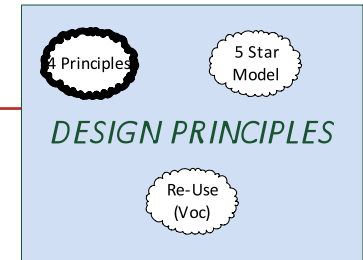


Platform Linked
Data Nederland

Linked Open Data: Design Principles



Defining linked data...



“Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens.”

EC ISA Case Study: How Linked Data is transforming eGovernment

The **four design principles** of Linked Data (*by Tim Berners Lee*):

Use Uniform Resource Identifiers (URIs) as names for things.

Use HTTP URIs so that people can look up those names.

When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).

Include links to other URIs so that they can discover more things.

See also:

http://www.youtube.com/watch?v=4x_xzT5eF5Q

<http://www.w3.org/DesignIssues/LinkedData.html>

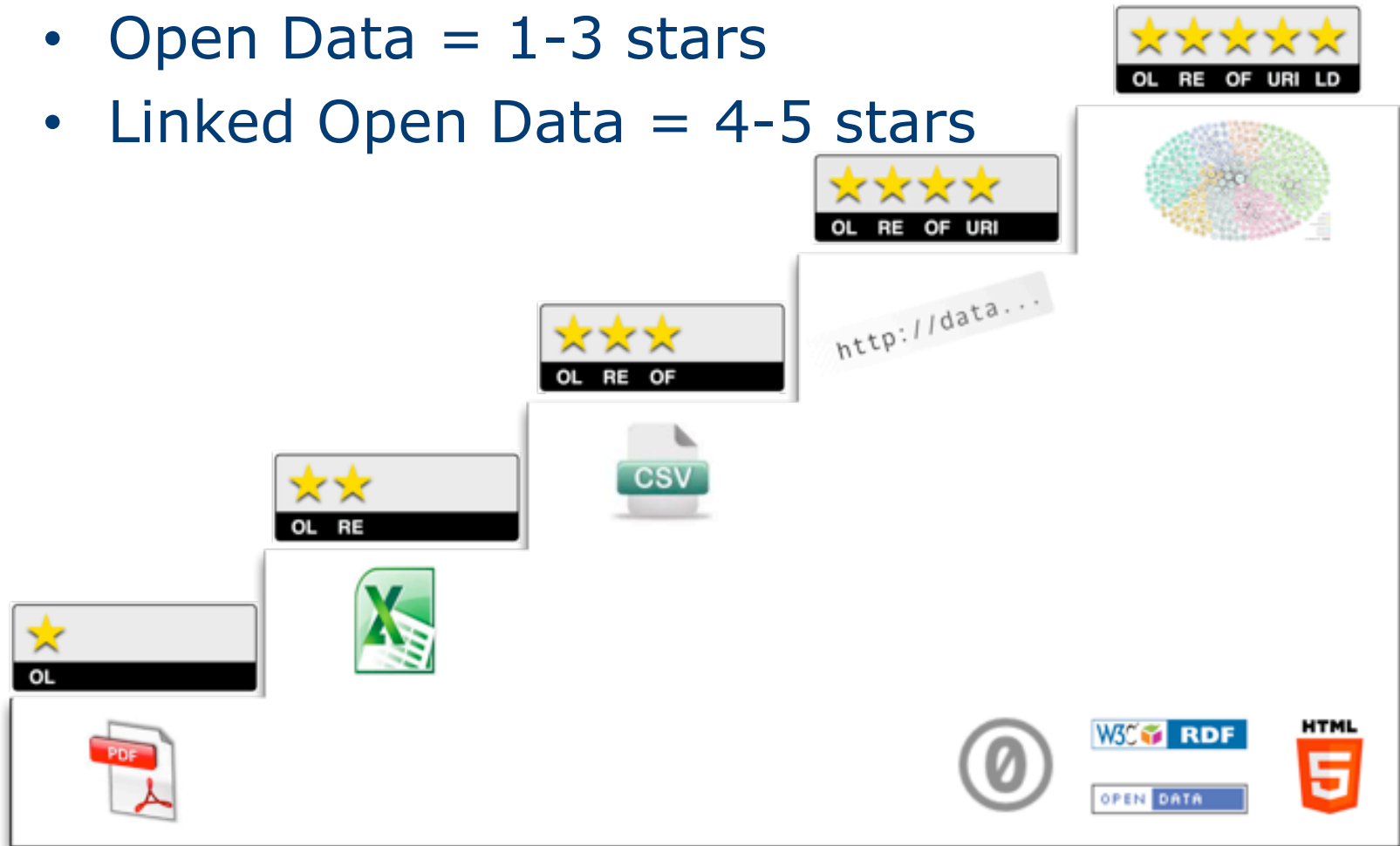
<http://www.youtube.com/watch?v=uju4wT9uBIA>



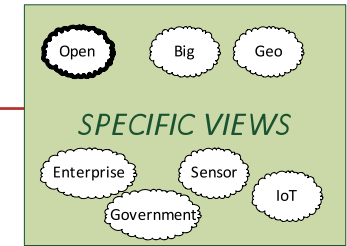
Platform Linked
Data Nederland

5 star model based on “W3C standards”

- Open Data = 1-3 stars
- Linked Open Data = 4-5 stars



Linked data vs. open data



“Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.”

- OpenDefinition.org

Open data

Data can be published and be publicly available under an open licence without linking to other data sources.



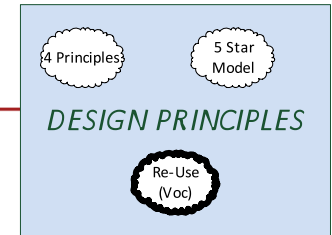
Linked data

Data can be linked to URIs from other data sources, using open standards such as RDF without being publicly available under an open licence.

See also:

Cobden et al., A research agenda for Linked Closed Data

http://ceur-ws.org/Vol-782/CobdenEtAl_COLD2011.pdf



Model your data – reuse if possible, mint if necessary

- **Reuse** existing vocabularies **as much as possible**.
 - If you determine there is no reusable, authoritative source for the specific domain, **create your own using**:
 - RDF Schema (RDFS): Basic RDF vocabulary to describe the classes and properties of classes.
 - Web Ontology Language (OWL): knowledge representation language for describing ontologies.

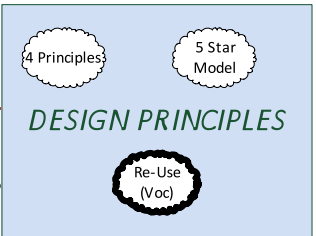
See also:

<http://www.slideshare.net/OpenDataSupport/model-your-data-metadata>

<http://www.w3.org/TR/owl-features/>

<http://www.w3.org/TR/rdf-schema/>

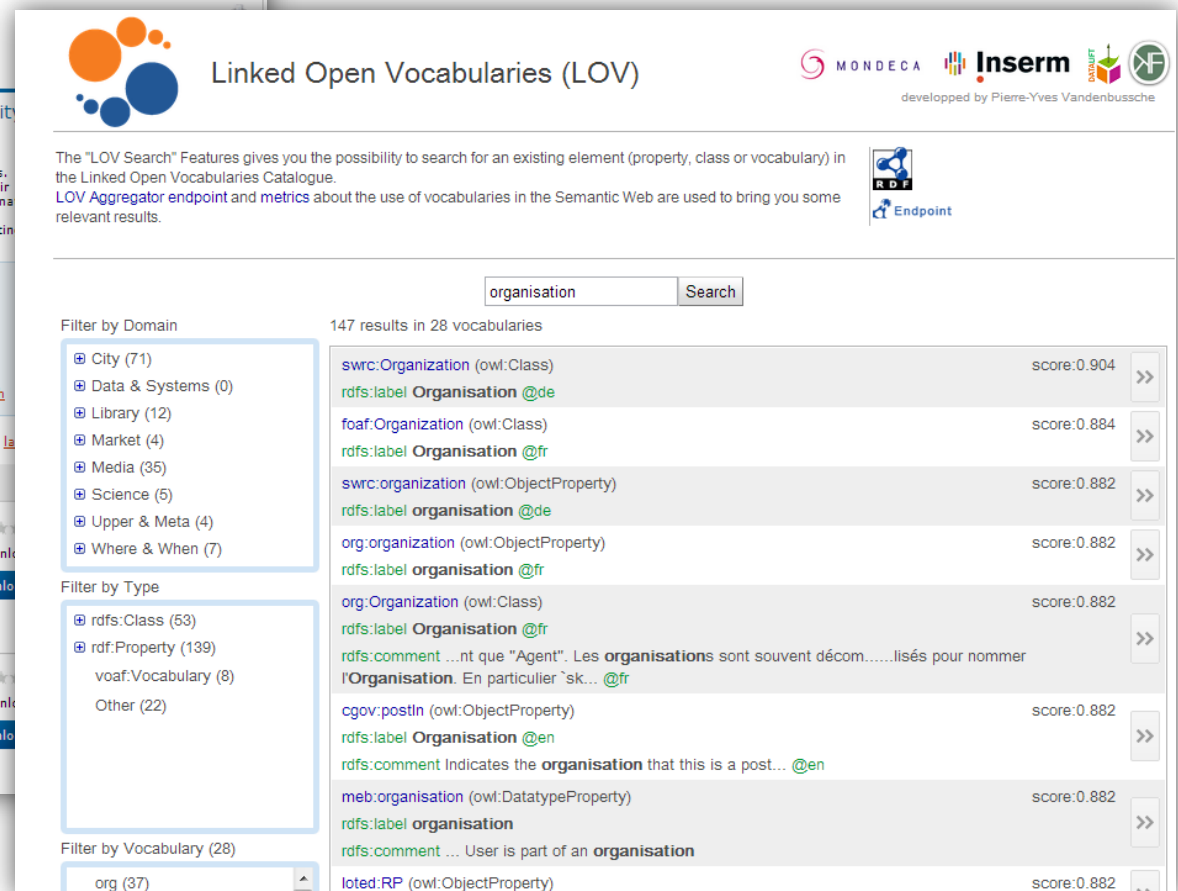
You can find reusable RDF vocabularies on..



<http://joinup.ec.europa.eu/>

 OPEN DATA SUPPORT

<http://lov.okfn.org/>



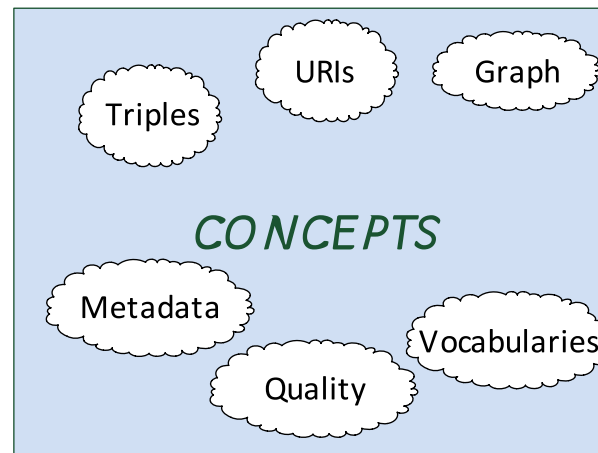
Slide 20





Platform Linked
Data Nederland

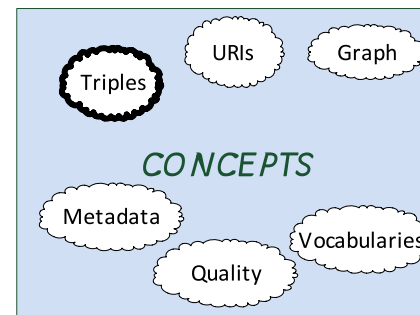
Linked Open Data: Concepts



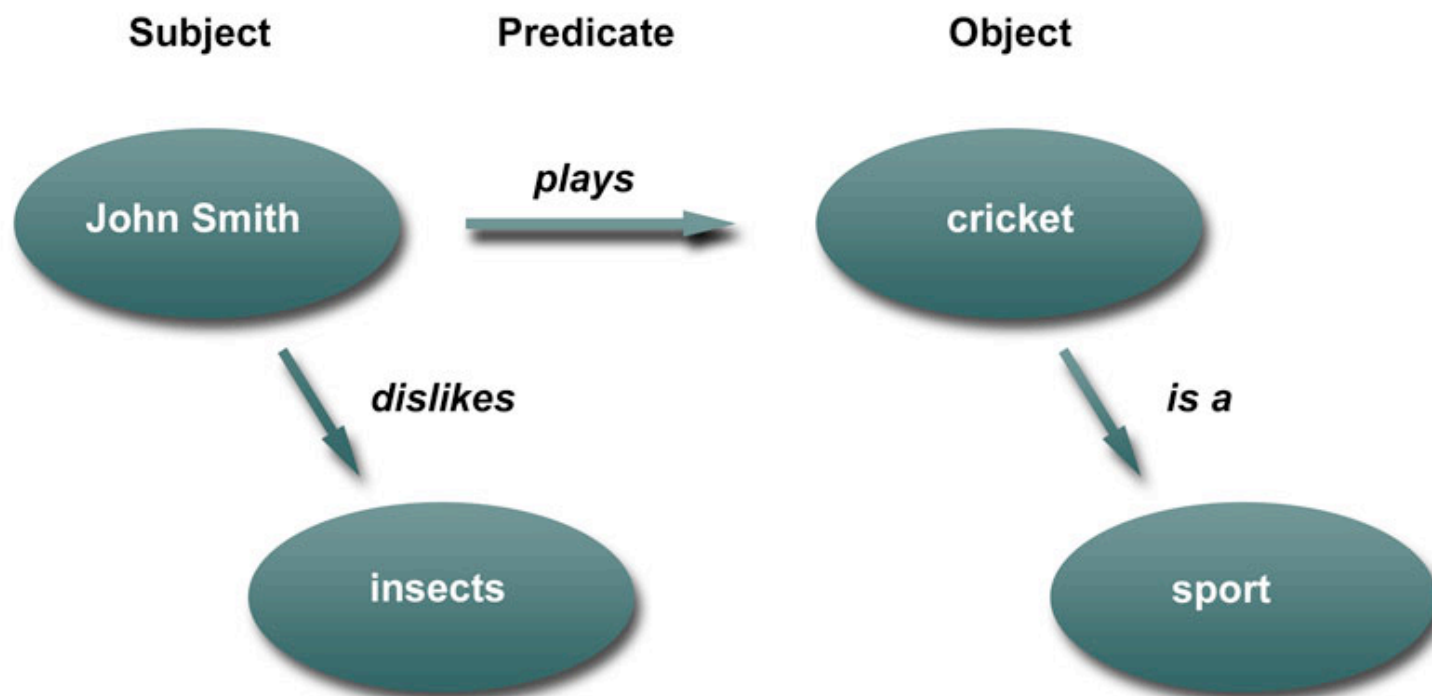


Platform Linked
Data Nederland

TRIPLES



- RDF – AAA principle

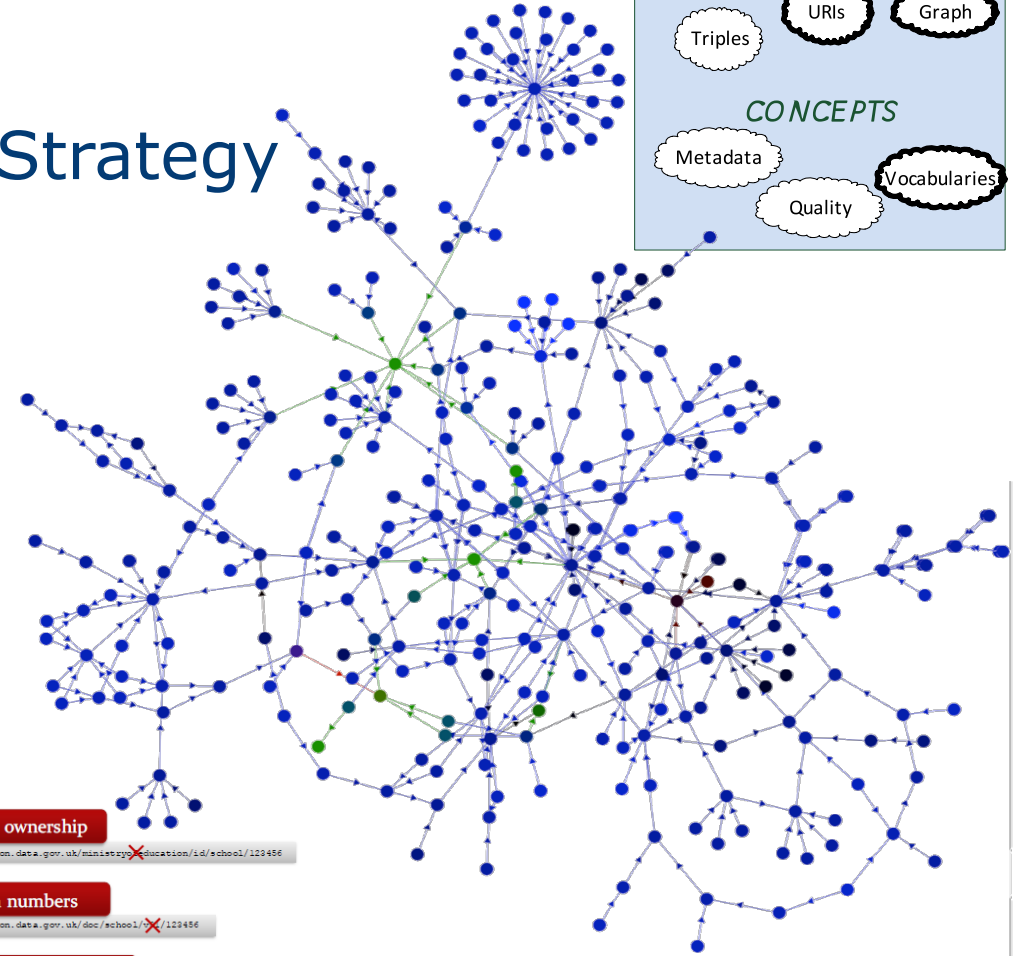
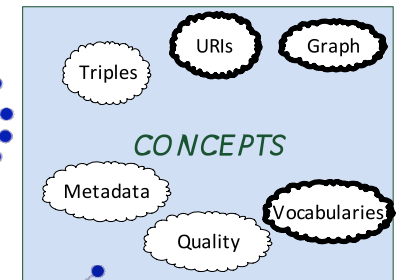




Platform Linked
Data Nederland

Linked (Open) Data Concepts

- URIs: National Strategy
- Vocabularies
- Graphs



Persistent URIs sets the foundations for Linked Data.



Follow the pattern

e.g. `http://(domain)/(type)/(concept)/(reference)`

Re-use existing identifiers

e.g. `http://education.data.gov.uk/id/school/123456`

Link multiple representations

e.g. `http://data.example.org/doc/foo/bar.html`

e.g. `http://data.example.org/doc/foo/bar.pdf`

Implement 303 redirects for real-world objects

e.g. `http://www.example.com/id/alice_brown`

Use a dedicated service

i.e. independent of the data originator

10
rules
for **persistent**
URIs



Avoid stating ownership

e.g. `http://education.data.gov.uk/ministry/education/id/school/123456`

Avoid version numbers

e.g. `http://education.data.gov.uk/doc/school/123456`

Avoid using auto-increment

e.g. `http://education.data.gov.uk/id/school1/123456`

e.g. `http://education.data.gov.uk/id/school1/123457`

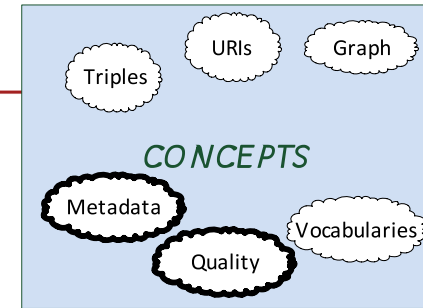
Avoid query strings

e.g. `http://education.data.gov.uk/doc/school1?id=123456`

Avoid file extensions

`http://education.data.gov.uk/doc/school1/123456.csv`

Cleansing your data & metadata



To ensure data and metadata can be published with an appropriate level of quality and minimum errors.

This means:

- Fixing errors.
- Transforming/homogenising formats.
- Aligning inconsistencies in data and metadata.
- Removing duplicate/redundant information.
- Adding lacking information.
- Making sure the information is up-to-date.

See also:

<http://www.slideshare.net/OpenDataSupport/introduction-to-rdf-sparql>

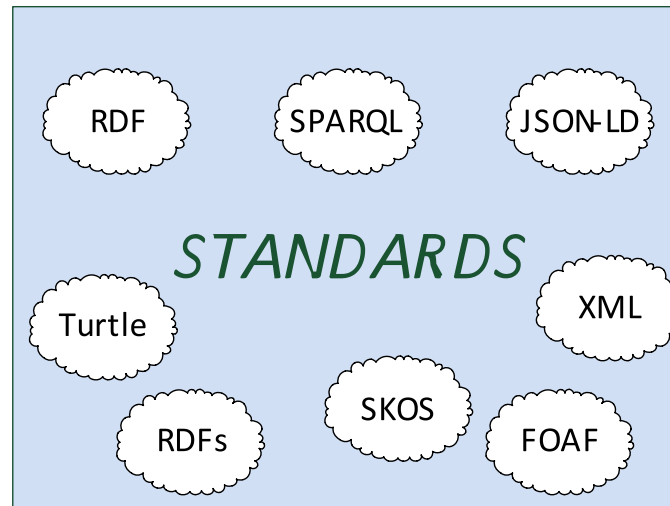
Cleanse your data with Open Refine (Google Refine) -

<https://code.google.com/p/google-refine/>



Platform Linked
Data Nederland

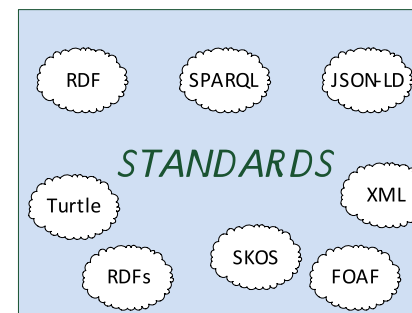
Linked Open Data: Standards





Platform Linked
Data Nederland

Standaarden



The Semantic Web Technology Stack (not a piece of cake...)

Most apps use only a subset of the stack

Querying allows fine-grained data access

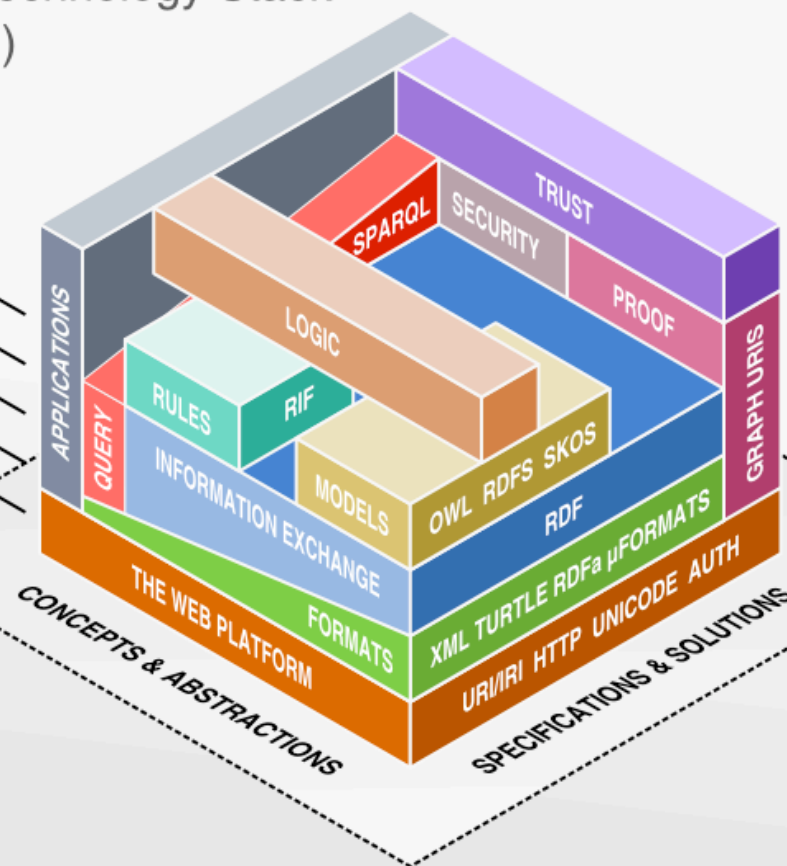
Standardized information exchange is key

Formats are necessary, but not too important

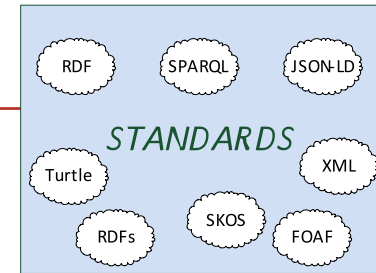
The Semantic Web is based on the Web

Linked Data uses a small
selection of technologies

LINKED DATA



RDF & SPARQL



The **Resource Description Framework** (RDF) is a syntax for representing data and resources in the Web

RDF breaks every piece of information down in **triples**:

- Subject – a resource, which may be identified with a URI.
- Predicate – a URI-identified reused specification of the relationship.
- Object – a resource or literal to which the subject is related.

<http://dbpedia.org/resource/Brussels> is the capital of “Belgium”.

OR

<http://dbpedia.org/resource/Brussels> is the capital of <http://dbpedia.org/resource/Belgium>.

Subject

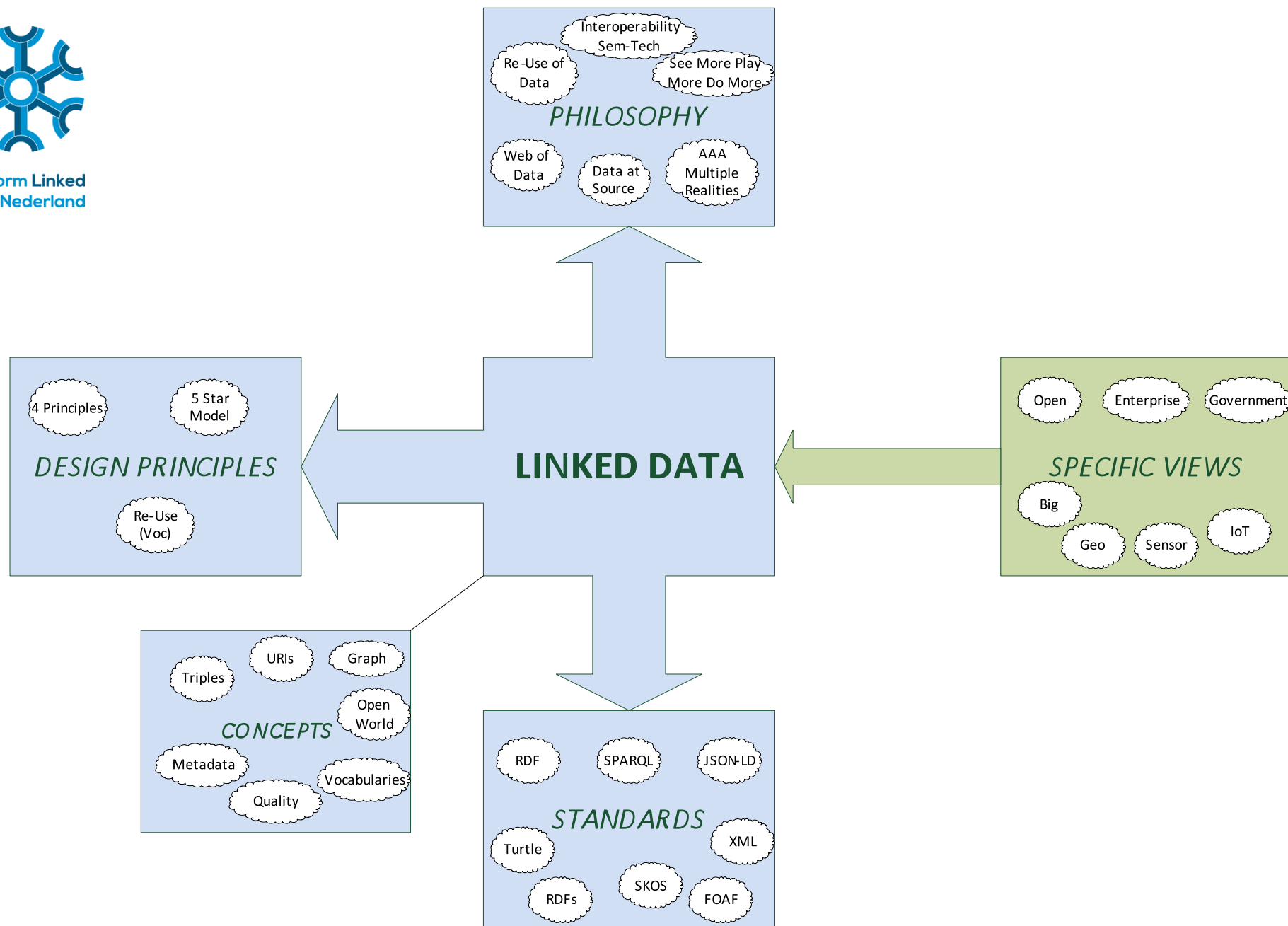
Predicate

Object

SPARQL is a standardised language for querying RDF data.



Platform Linked
Data Nederland





Platform Linked
Data Nederland

Linked (Open) Data – The approach

- Making your data available in a **structured, comprehensible** and **machine-readable** way.
- **Reusing** what already exists in terms of vocabularies and reference data.
- Reaching the right quality level by **cleansing** your data.
- Providing **licensing information** so that data consumers know what the conditions of reuse are.
- Providing a rich description (**metadata**).
- Using **semantic technologies** (RDF, HTTP URIs...) for describing your data.





Platform Linked
Data Nederland

LOD in practice

- Pharma (e.g. OpenPhacts)
- Health Care
- Libraries, Museums, Archiving
- Education, Universities
- Geo (Smart Cities)
- Media, Publishing (e.g. BBC, Wolters Kluwer)
- High Tech: (e.g. NXP)
- Logistics
- Government (Stelsel van Basisregistraties)

- Join the Platform Linked Data NL meetings...





Platform Linked
Data Nederland

Test your knowledge: <https://testmoz.com/185946>

Linked (Open) Data The Summary

Breaking down the walls of the silos in order to create more value.

- Allows for flexible integration of datasets from different sources, without needing the data to be moved.
- Fosters the reuse of information from reference/authoritative sources.
- Caters for assigning common identifiers in the form of HTTP URIs to things (e.g. people, products, business, locations...).
- Provides context to data – richer and more expressive data.
- The use of standard Web interfaces (such as HTTP and SPARQL) can simplify the use of data for machines.

“De echte waarde ligt in het combineren van data, bijvoorbeeld statistische data met tweets over een wijk.”

*Constantijn van Oranje,
Kabinetschef van EU-
commissaris Kroes*

Needed for societal and economic impact with (open) data

But not that easy....

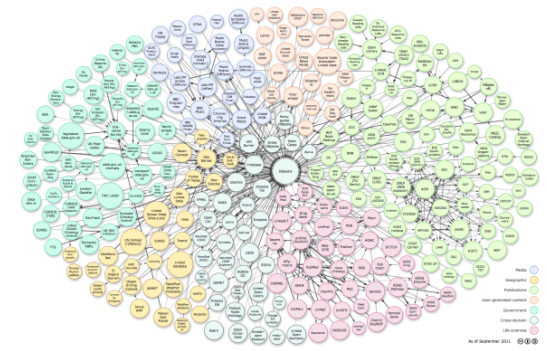
- Limited Knowledge & Tools
- Triple Explosion
- Precision (same as)
- B2B transaction world
- ...





Platform Linked
Data Nederland

What is Linked Data?



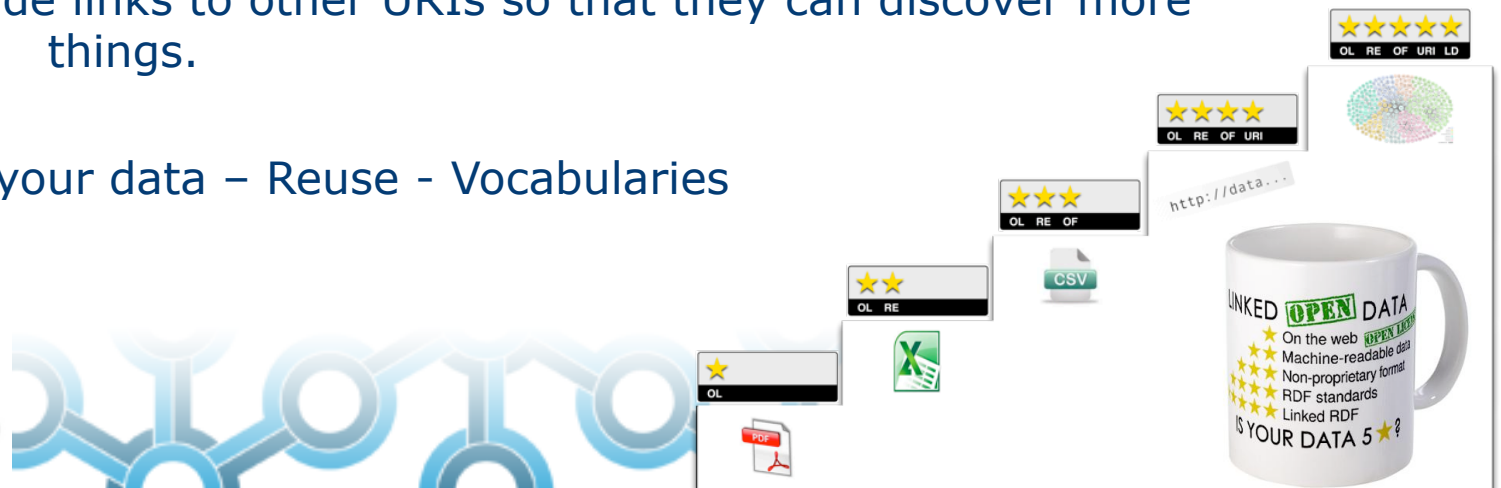
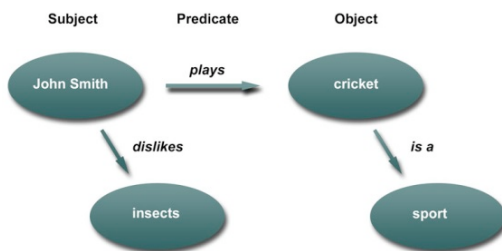
"Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens."



The **four design principles** of Linked Data (*by Tim Berners Lee*):

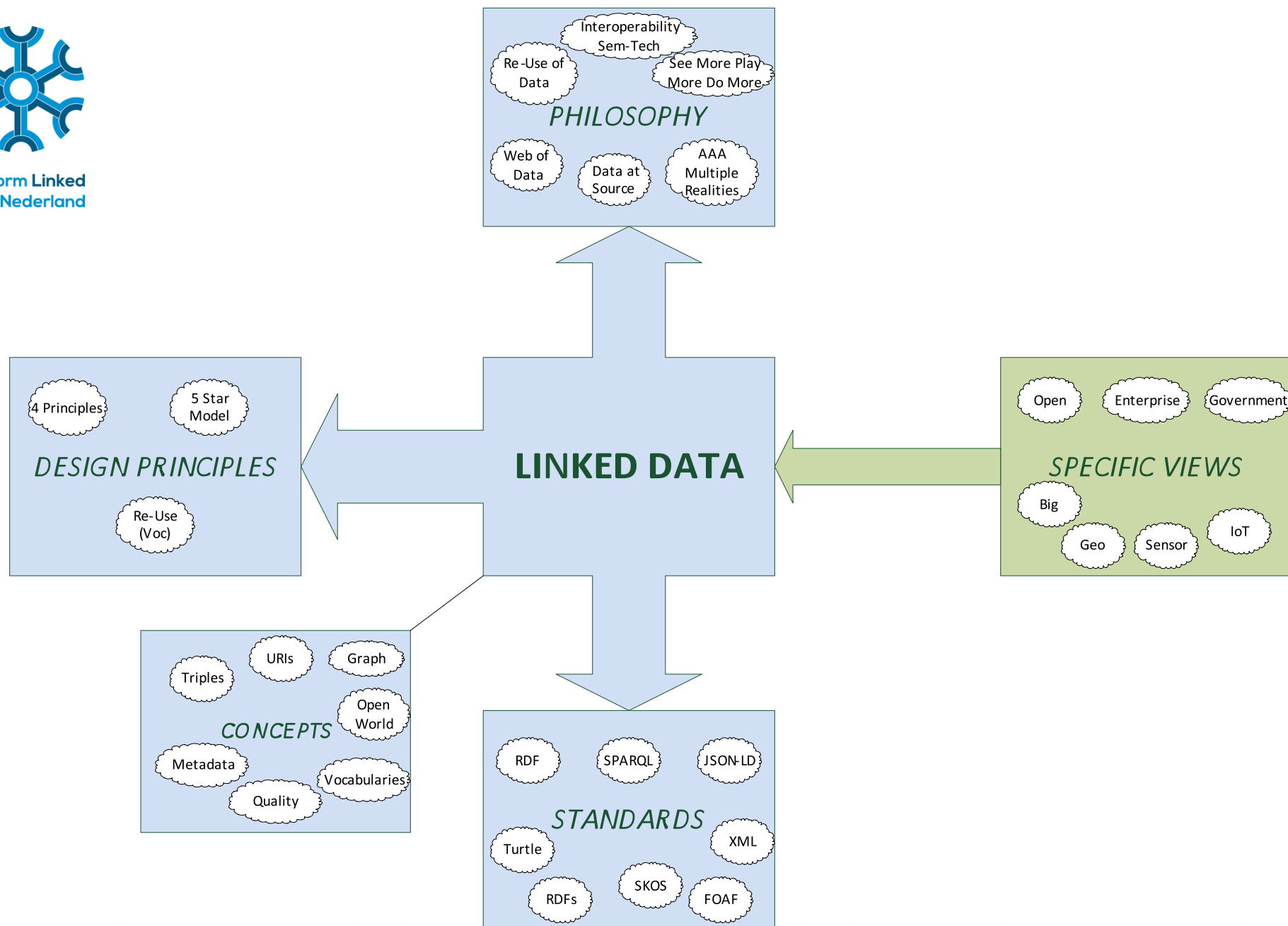
1. Use Uniform Resource Identifiers (URIs) as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs so that they can discover more things.

Model your data – Reuse – Vocabularies





Platform Linked
Data Nederland



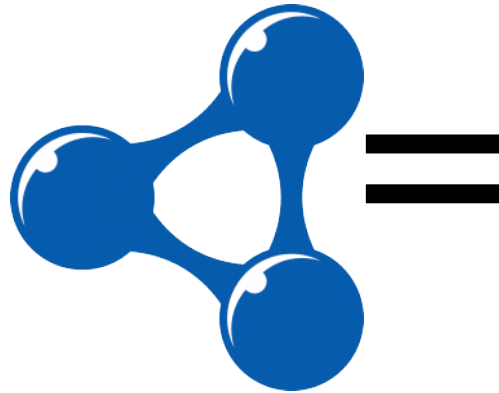


Platform Linked
Data Nederland

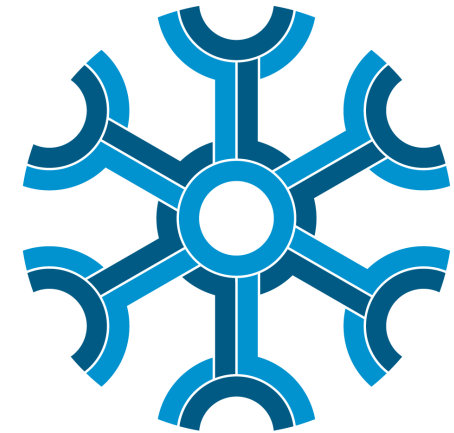
Open Innovation + Linked Data = Platform Linked Data Netherlands



+



=



Platform Linked
Data Nederland

Open Community for Linked Data knowledge exchange





2014-2017

PLATFORM LINKED DATA NEDERLAND

CONTACT: ERWIN FOLMER (ERWIN.FOLMER@TNO.NL)

